The Practical Limits of Photon Communication

R. J. McEliece, E. R. Rodemich, and A. L. Rubin
Communications Systems Research Section

We show that for photon communication, the rate R_0 = 1 nat per photon is the rate beyond which one encounters an explosive increase in both the required ratio of peak-to-average signal power and in the required bandwidth expansion. On the basis of these results we conjecture that no practical photon communication system can be designed to operate above 10 nats per photon.

I. Introduction

In a recent report (Ref. 1), it was shown that the R_0-parameter associated with noiseless optical communication using photon-counting techniques (hereafter we call this "photon communication") is one nat per photon. Now for any channel, R_0 is widely believed to be the rate above which the implementation of a reliable communication system becomes very difficult, but there is no really sound mathematical support for this belief. In this paper, however, we will give rigorous mathematical substantiation to this "R_0 conjecture," for the special case of photon communication.

Roughly speaking, we shall prove that for photon communication, the rate R_0 = 1 is the rate beyond which one encounters an explosive increase in both the required ratio of peak-to-average signal power and in the required bandwidth expansion.

Precisely speaking, what we shall prove is this. Let ρ denote the rate (in nats per photon) of a given reliable photon communication system, let α denote its required ratio of peak-to-average signal power, and let β denote its required bandwidth expansion factor. Then necessarily, as we will show in Sections II and III,

$$\alpha \geqslant e^{\rho-1} - 1 \tag{1}$$

$$\beta \geqslant \frac{e^{\rho-1} - 1}{\rho} \tag{2}$$

Thus as ρ increases linearly, both α and β must increase exponentially. On the basis of Eqs. (1) and (2), we conjecture that no practical photon communication system can be designed with $\rho \geqslant 10$. On the other hand, in Ref. 2 it was shown that one could design a practical system at about $\rho = 3$ using pulse position modulation and Reed-Solomon codes. Thus, the gap between what is presently practical and what may someday be practical is reasonably small. This is in spite of the fact that channel capacity (i.e., the largest ρ that is theoretically possible) is extremely large (Ref. 3).

II. The Poisson Channel Model

We assume that any photon communication system works as follows. The time interval during which communication

takes place is divided into many subintervals ("slots"), each of duration t_0 seconds. The transmitter is a laser which is pulsed during each time slot; it may be pulsed with a different intensity in each slot. At the receiver is a photon counter, which accurately counts the number of photons received during each time slot. We denote by x_i the expected number of photons received during the ith time slot; x_i will be called the *intensity* of the ith pulse.

It may be that "noise photons" are present in such a system, but in many cases of practical interest, noise photons are extremely rare. (For example, in a careful analysis of a potentially practical system, Katz estimated the rate of arrival of noise photons to be around 10^{-3} per second.) In any event we shall make the assumption that no noise photons exist. In this case, because of the Poisson nature of photon arrivals, the probability that exactly k photons will be received during a slot in which the laser was pulsed with intensity x is $e^{-x}x^k/k!$

Thus described, the optical channel is a discrete memoryless channel with input alphabet equal to the set of nonnegative real numbers (the possible values for the intensities x_i), and output alphabet equal to the set of nonnegative integers (the possible outputs of the photon counter). If a real number x is transmitted, the probability that the integer k will be received is given by

$$p(\mathrm{k}|x) = e^{-x}\frac{x^k}{k!} \tag{3}$$

We call the channel described by Eq. (3) the *Poisson channel.*

A code for this channel is a set of vectors $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{in})$, $i = 1, \ldots, M$, of length n. Each component x_{ij} is a nonnegative real number and represents an intensity of the transmitting laser. Assuming that each component of a code word requires one time slot for transmission, the rate of such a code is

$$R = \frac{\log M}{n} \text{ nats per slot} \tag{4}$$

On the other hand, each component x_{ij} represents an average number of (received) photons, and so the code's rate in nats per photon is

$$\rho = \frac{R}{\mu} \text{ nats per photon, where} \tag{5}$$

$$\mu = \left(\sum_{i,j} x_{ij}\right) \Big/ nM, \text{ photons per slot (average)} \tag{6}$$

The rate R in Eq. (4) is a measure of "bandwidth expansion." If, for example, we are transmitting at a rate of A nats per second, using a code of rate R nats per slot, it follows that we require A/R slots per second. Thus, the slot rate is equal to the nat rate multiplied by the factor $1/R$. We thus define

$$\beta = \frac{n}{\log M} \text{ slots per nat} \tag{7}$$

and call β the *bandwidth expansion factor.*

In Eq. (6), μ represents the *average* pulse intensity, i.e. the average number of photons received per time slot. At a given frequency this number is proportional to the *average* received power. On the other hand, the quantity

$$L = \max_{i,j} x_{ij} \tag{8}$$

represents the *maximum* required pulse intensity required by this particular code, and it is proportional to the maximum or *peak* received power. We denote the ratio of L to μ by α:

$$\alpha = \frac{L}{\mu} = \frac{\max_{i,j}(x_{ij})}{\frac{1}{n \cdot M}\sum_{i,j} x_{ij}} \tag{9}$$

We have now precisely defined the quantities α, β, and ρ mentioned in Section I. Our proof of inequalities in Eqs. (1) and (2) rests on certain information – theoretic results about the Poisson channel, which we now describe.

We would like to compute the information – theoretic capacity of the Poisson channel, whose statistics are given in Eq. (3); that is, the maximum possible mutual information $I(X;K)$, where X is a nonnegative random variable, and K is a nonnegative integer-valued random variable related to X by the conditional probabilities. However, it is very easy to see that this maximum is infinite (take X to be a discrete random variable which assumes a very large number of values which are very far apart). To get a meaningful problem, we must restrict X somehow. The most natural restriction (Eq. (6)) is to fix the *mean* of X; and so we define

$$C(\mu) = \sup\,[I(X;K): E(X) = \mu] \tag{10}$$

According to Shannon's noisy-channel coding theorem (see Ref. 4, Chapter 7), $C(\mu)$ represents the maximum possible rate (in nats per channel use) of a reliable photon communication system which is restricted to operate at an average of μ photons per slot or less.

A second possible restriction is on the *maximum* value that X can assume (Eq. (8)). Thus we define

$$C(\mu,L) = \sup [I(X;K): E(X) = \mu, X \leqslant L] \tag{11}$$

Again, according to Shannon's theorem, $C(\mu,L)$ represents the maximum possible rate for any photon communication system which is restricted to operate at an average of $\leqslant \mu$ photons per slot, and a maximum of L photons per slot.

In Section III, we will prove the following results about $C(\mu)$ and $C(\mu,L)$:

$$C(\mu) \leqslant \log(1+\mu) + \mu \log\left(1+\frac{1}{\mu}\right) \tag{12}$$

$$C(\mu,L) \leqslant \mu \log \frac{L}{\mu}, \text{ if } L \leqslant 1 \tag{13}$$

In the remainder of this section, we will show how Eqs. (12) and (13) can be used to prove our main results, Eqs. (1) and (2).

First, note that by the converse to the noisy-channel coding theorem, the rate R of a reliable communication system which operates at an average of μ photons per slot is bounded by

$$R < C(\mu) \tag{14}$$

Now $C(\mu)$ itself is bounded by Eq. (12), and since $\log(1+\mu) < \mu$, we have

$$R < \mu\left[1 + \log\left(1+\frac{1}{\mu}\right)\right] \tag{15}$$

The rate of this system measured in nats per photon is, by Eq. (5), R/μ, and so

$$\rho < 1 + \log\left(1+\frac{1}{\mu}\right) \tag{16}$$

for $\rho > 1$ a simple algebraic manipulation of Eq. (16) yields

$$\mu < (e^{\rho-1} - 1)^{-1} \tag{17}$$

Now since the bound on the right of Eq. (15) is easily seen to be an increasing function of μ, it follows from Eqs. (15) and (17) that

$$R < \frac{\rho}{e^{\rho-1} - 1} \tag{18}$$

but $R = \beta^{-1}$. This proves Eq. (2).

To prove Eq. (1) observe that Eq. (13) tells us that $\alpha \geqslant e^{\rho}$ for $L \leqslant 1$. This is stronger than Eq. (1); thus Eq. (1) can only fail for $L > 1$. But if $L > 1$, the ratio $\alpha = L/\mu$ is by Eq. (17) greater than $e^{\rho-1} - 1$. This proves Eq. (1).

It now remains to prove Eqs. (12) and (13). This we do in the next section.

III. Proof of Eqs. (12) and (13)

In this section we shall give proofs of the important inequalities of Eqs. (12) and (13). We begin with Eq. (12).

By definition, $C(\mu)$ is the largest possible value which can be assumed by the mutual information $I(X;K)$ when the test source X is restricted to satisfy $E(X) = \mu$. But by well-known results on mutual information (see Ref. 5, Chapter 1),

$$\begin{aligned} I(X;K) &= H(K) - H(K|X) \\ &\leqslant H(K) \end{aligned} \tag{19}$$

where in Eq. (19) $H(K)$ denotes the entropy $\Sigma_k p_k \log p_k^{-1}$ of the random variable K. Since for the Poisson channel, $E(K|X) = X$, it follows that $E(K) = E[E(K|X)] = E(X)$, and so the random variable K has the same mean as X, viz., μ.

The problem of maximizing the entropy of a nonnegative integer-valued random variable with a given mean is easily handled with standard information-theoretic techniques. Omitting the details (see Ref. 5, Problem 1.8), the result is

$$H(K) \leqslant \log(1+\mu) + \mu \log\left(1+\frac{1}{\mu}\right) \tag{20}$$

(provided $E(K) = \mu$, K assumes only nonnegative integer values). Equations (19) and (20) combine to give Eq. (12).

We turn now to Eq. (13), which lies somewhat deeper. The problem is to maximize $I(X;K)$ given that the distribution of X is restricted to $[0,L]$ and has mean μ.

We introduce the notation (see Eq. (13))

$$p_k(x) = e^{-x}\frac{x^k}{k!} \tag{21}$$

to denote the probability that k will be received given that x is transmitted. We define

$$\overline{p}_k = E[p_k(X)] \tag{22}$$

the expectation being with respect to the distribution of X. The quantity $\overline{p}_k$ represents the probability that k will be the channel's output, given that the input is the random variable X.

Using standard techniques of convex analysis, it is now possible to show that a particular distribution confined to $[0,L]$ with expection μ maximizes $I(X;K)$ if and only if for some constants C and λ,

$$F(x) = \sum_{k=0}^{\infty} p_k(x)\log\frac{p_k(x)}{\overline{p}_k} + \lambda(x-\mu) - C \leqslant 0,\ 0 \leqslant x \leqslant L \tag{23}$$

where equality holds in Eq. (23) at all points of support of the distribution. (Eq. (23) is essentially the same as Theorem 4.5.1 in Gallager (Ref. 4). The only differences are that our channel has a countably infinite alphabet, rather than a finite one, and that we have an extra constraint $E(X) = \mu$, which necessitates the Lagrange multiplier term $\lambda(x-\mu)$. However, the modification of Gallager's analysis needed to arrive at Eq. (23) is quite easy, and we omit it.)

Equation (23) is a very strong condition that must be satisfied by an extremel distribution. For example, we use it to show the following.

Lemma: For any L,μ, a maximizing distribution can have mass at at most one point on (0,1).

Proof: Define

$$g_k = k!\,\overline{p}_k = E(e^{-X}X^k) \tag{24}$$

Then by Schwarz's inequality $[E(X_1X_2)]^2 \leqslant E(X_1^2)\,E(X_2^2)$ applied to $X_1 = e^{-X/2}\,X^{(k-1)/2}$, $X_2 = e^{-X/2}\,X^{(k+1)/2}$, we have

$$g_k^2 \leqslant g_{k-1}\,g_{k+1}, \quad k \geqslant 1 \tag{25}$$

The function $F(x)$ of Eq. (23) can be written in the form

$$F(x) = x(\log x - 1) + \sum_{k=0}^{\infty} p_k(x)\log\frac{1}{g_k} + \lambda(x-\mu) - C \tag{26}$$

If we differentiate $F(x)$ twice, using the formula $p_k'(x) = p_{k-1}(x) - p_k(x)$, we get

$$F''(x) = \frac{1}{x} + \sum_k [p_{k-2}(x) - 2p_{k-1}(x) + p_k(x)]\log\frac{1}{g_k}$$

$$= \frac{1}{x} + \sum_k p_k(x)\log\frac{g_{k+1}^2}{g_k g_{k+2}} \tag{27}$$

$$e^xF''(x) = \frac{e^x}{x} - \sum_{k=0}^{\infty}\frac{x^k}{k!}\log\frac{g_k g_{k+2}}{g_{k+1}^2} \tag{28}$$

Each coefficient in the series Eq. (28) is by Eq. (25) non-negative. Since e^x/x is a decreasing function for $0 < x < 1$, it follows that $e^xF''(x)$ is decreasing in this range also. Since $e^xF''(x)$ is positive at $x = 0+$, $e^xF''(x)$, and so $F''(x)$ also, can be zero for at most one value of $0 < x < 1$.

An extremel distribution by Eq. (23) must have $F(x) \leqslant 0$ for all $0 < x < L$. Since $F(x)$ is differentiable, it follows that $F'(x) = 0$ whenever $F(x) = 0$. Now if say $F(x_1) = F(x_2) = 0$ with $0 < x_1 < x_2 < 1$, then there exists $x_1 < x_3 < x_2$ with $F'(x_3) = 0$. By the above remarks $F'(x_1) = F'(x_2) = 0$ as well. This in turn implies the existence of x_4, x_5: $x_1 < x_4 < x_3 < x_5 < x_2$ with $F''(x_4) = F''(x_5) = 0$. But we argued above that $F''(x)$ could vanish at most once on (0,1). Thus, $F(x)$ can vanish at most once on (0,1), i.e., the optimizing distribution can have mass, at most, at one point in (0,1). This completes the proof of the lemma.

We now use the lemma to prove Eq. (13). Since $L \leqslant 1$, the lemma tells us that a distribution on $[0,L]$ with $E(x) = \mu$ can have mass only at $x = 0$ and one other point $x = \ell \leqslant L$. We shall complete the proof of Eq. (13) by showing that for any distribution concentrated at $x = 0$ and $x = L$ the resulting mutual information $I(X;K)$ satisfies

$$I(X;K) \leqslant \mu\log\frac{L}{\mu} \tag{29}$$

(It is sufficient to take $\ell = L$ because the right side of Eq. (29) is an increasing function of L.)

Thus, let X have distribution

$$P(X = 0) = p$$
$$P(X = L) = q, \quad p + q = 1 \tag{30}$$

Then a straightforward calculation yields

$$I(X;K) = (1 - Q) \log \frac{1}{1 - Q} + Q \log \frac{1}{q} - L(q - Q)$$

$$Q = q(1 - e^{-L}) \tag{31}$$

Our goal is to show that this quantity is $\leqslant \mu \log (L/\mu)$; but $\mu = E(X) = qL$. So we must show that the right side of Eq. (31) is $\leqslant qL \log 1/q$. Subtracting this quantity from Eq. (31), we define for a fixed $L > 0$

$$F(q) = (1 - Q) \log \frac{1}{1 - Q} + (Q - qL) \log \frac{1}{q} - L(q - Q)$$

and wish to show that $f(q) \leqslant 0$ for $0 \leqslant q \leqslant 1$. This is easily seen, given the following (whose straightforward verifications are omitted):

$$f(0) = f(1) = 0 \tag{32}$$

$$f'(0) = -\infty, f'(1) = 0 \tag{33}$$

$$f''(q_0) = 0 \text{ only for } q_0 = \frac{1}{1 - e^{-L}} - \frac{1}{L} \tag{34}$$

Because $f(0) = 0$, $f'(0) = -\infty$, $f(q)$ is negative for all sufficiently small q. If now $f(q) = 0$ for $0 < q < 1$, it would necessarily follow that f' would vanish at two interior points of (0,1). Since also $f'(1) = 0, f''$ would vanish at two points of (0,1), contradicting Eq. (34). Thus $f(q) \leqslant 0$ for all $0 \leqslant q \leqslant 1$, and this completes the proof of Eq. (13).

References

1. McEliece, R. J., "The R_0-Parameter for Optical Communication Using Photon Counting," *DSN Progress Report 42-53*, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1979, pp. 62-65.
2. McEliece, R. J. and L. R. Welch, "Coding for Optical Channels With Photon Counting," *DSN Progress Report 42-52*, Jet Propulsion Laboratory, Pasadena, Calif., Aug. 15, 1979, pp. 61-66.
3. Pierce, J. R., "Optical Channels: Practical Limits With Photon Counting," *IEEE Trans. Commun.*, COM-26 (1978), pp. 1819-1821.
4. Gallager, R. G., *Information Theory and Reliable Communication*. Wiley and Sons, New York, 1968.
5. McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley, Reading, Mass., 1977.